Systematic Study of Selected Diagonalization Methods for Configuration Interaction Matrices

MATTHEW L. LEININGER,[1] C. DAVID SHERRILL,[2] WESLEY D. ALLEN,[3] HENRY F. SCHAEFER III[3]

[1]*Sandia National Laboratory, 7011 East Ave., M. S. 9217, Livermore, California 94551-0969*
[2]*School of Chemistry and Biochemistry, Georgia Institute of Technology, Atlanta, Georgia 30332-0400*
[3]*Center for Computational Quantum Chemistry, University of Georgia, Athens, Georgia 30602-2556*

Received 23 February 2001; accepted 23 April 2001

Dedicated to Professor Paul von R. Schleyer

ABSTRACT: Several modifications to the Davidson algorithm are systematically explored to establish their performance for an assortment of configuration interaction (CI) computations. The combination of a generalized Davidson method, a periodic two-vector subspace collapse, and a blocked Davidson approach for multiple roots is determined to retain the convergence characteristics of the full subspace method. This approach permits the efficient computation of wave functions for large-scale CI matrices by eliminating the need to ever store more than three expansion vectors ($\mathbf{b}_i$) and associated matrix-vector products ($\boldsymbol{\sigma}_i$), thereby dramatically reducing the I/O requirements relative to the full subspace scheme. The minimal-storage, single-vector method of Olsen is found to be a reasonable alternative for obtaining energies of well-behaved systems to within μE_h accuracy, although it typically requires around 50% more iterations and at times is too inefficient to yield high accuracy (ca. 10^{-10} E_h) for very large CI problems. Several approximations to the diagonal elements of the CI Hamiltonian matrix are found to allow simple on-the-fly computation of the preconditioning matrix, to maintain the spin symmetry of the determinant-based wave function, and to preserve the convergence characteristics of the diagonalization procedure.
J Comput Chem 22: 1574–1589, 2001

Keywords: configuration interaction; diagonalization methods; Davidson method; *ab inito* methods; symmetric eigenvalue problem

Correspondence to: W. D. Allen; e-mail: wdallen@ccqc.uga.edu
Contract/grant sponsor: U.S. National Science Foundation; contract/grant number: CHE-9815397

Introduction

The configuration interaction (CI) approach dates back to the earliest days of quantum mechanics and still endures as one of the most flexible techniques in computational chemistry for obtaining electronic wavefunctions of the ground and excited states of atomic and molecular systems.[1–3] The CI method involves expressing the Schrödinger equation as a matrix eigenvalue problem [eq. (1)] within a finite-dimensional n-particle basis of Slater determinants, or alternatively a spin-adapted basis of configuration state functions.

$$\mathbf{H}\mathbf{c}^k = \lambda^k \mathbf{c}^k. \tag{1}$$

Matrix elements of the Hamiltonian operator are written in terms of standard one- and two-electron molecular integrals and coupling coefficients between the n-particle basis functions.[3–9] This numerical representation of the Hamiltonian leads to matrices whose dimensions grow exponentially with respect to the one-particle basis and the number of electrons. The standard methods[10, 11] for the determination of all eigenvalues and eigenvectors of real, symmetric $N \times N$ matrices necessitate greater than N^2 storage elements and scale as $\mathcal{O}(N^3)$.

For CI computations involving 10^6 Slater determinants, the standard methods for symmetric matrices would require about 16 terabytes of storage space, and currently take many years to complete. A more efficient approach is to use iterative techniques that exploit the sparsity of the Hamiltonian, avoid the explicit storage of the Hamiltonian matrix by forming matrix–vector products directly, and only solve for the lowest single or few eigenvalues and eigenvectors.[4, 5, 7, 9, 12–14] Such techniques have produced benchmark CI computations involving over 10^9 Slater determinants.[15–20]

The Davidson algorithm[12] is the traditional large-scale, iterative diagonalization method of computational quantum chemistry, being a particularly effective scheme among the host of subspace methods for extracting selected eigenvectors.[11, 13, 14, 21–25] This technique typically convergences in 10–20 iterations and requires the storage of an expansion vector ($\mathbf{b}_i$) and a matrix-vector product ($\boldsymbol{\sigma}_i = \mathbf{H}\mathbf{b}_i$) per desired root per iteration. For CI computations involving multiple roots, near degeneracies, or more than several million determinants, the Davidson procedure may suffer from insufficient disk storage, I/O delays, and slow convergence. The disk space bottleneck may be relieved by a periodic collapse of the expansion space, in which the expansion vectors are replaced by one or a few of the best current approximations for each eigenvector sought; however, this approach reduces the size of the variational expansion space, and thus has the potential to impede convergence. The methods of Olsen[15] and Mitrushenkov[26] diminish the formidable storage requirements in this manner in favor of larger numbers of iterations. Numerous of iterative schemes have been proposed in recent years that offer efficacious alternatives or modifications to the traditional Davidson method. In this letter we scrutinize many of these proposals, and systematically test their efficiency on a number of electronic structure problems, specifically, various CI matrices of H_2O, O_3, N_2, C_2, C_2H_4, SO_2, HCN, SiO, C_2H_2, NH_3, and F_2. We include consideration of the advantages of approximating the diagonal elements of the Hamiltonian matrix to allow efficient on-the-fly evaluation of the Davidson preconditioning matrix while maintaining spin symmetry in the determinantal expansion space. On the basis of these numerous model computations, recommendations are then drawn that mesh the various proposals into maximally efficient diagonalization algorithms within the limits of available computational resources.

Methods

DAVIDSON'S METHOD

The Davidson diagonalization method,[12] also called the Davidson–Liu or simultaneous expansion method,[27] is classified as a subspace iterative approach in which the eigenvectors of interest are expanded in a linear orthonormal vector space,

$$\mathbf{x}^k = \sum_i^L \alpha_i^k \mathbf{b}_i, \tag{2}$$

where the α_i^k are expansion coefficients, $\mathbf{b}_i$ is the ith expansion vector, and $\mathbf{x}^k$ is the current iteration's approximation to $\mathbf{c}^k$, the kth eigenvector of the Hamiltonian. A subspace representation of the Hamiltonian matrix is generated with the expansion vectors and matrix–vector products of the Hamiltonian matrix with the expansion vectors,

$$G_{ij} = (\mathbf{b}_i, \mathbf{H}\mathbf{b}_j) = (\mathbf{b}_i, \boldsymbol{\sigma}_j), \qquad 1 \le i, j \le L, \tag{3}$$

where $\boldsymbol{\sigma}_j = \mathbf{H}\mathbf{b}_j$ and the expansion coefficients are subsequently determined as the eigenvectors of $\mathbf{G}$:

$$\mathbf{G}\boldsymbol{\alpha}^k = \rho^k \boldsymbol{\alpha}^k. \tag{4}$$

The subspace matrix $\mathbf{G}$ is easily diagonalized by standard methods, because the number of expan-

sion vectors is much smaller than the size of the CI Hamiltonian matrix. As the number of expansion vectors is appropriately increased, the CI eigenvector approximated by eq. (2) will approach the exact eigenvector, and the eigenvalues (ρ^k) of the small subspace matrix **G** will approach those of the CI Hamiltonian matrix (λ^k). This method enables the iterative diagonalization of large-scale matrices through the application of standard diagonalization methods to a small subspace eigenvalue problem. The most CPU demanding step involves the formation of the $\boldsymbol{\sigma}$ vectors; however, for large CI problems that exhaust available memory, the algorithm easily becomes I/O bound.

An expression for the Davidson expansion vectors ($\mathbf{b}_i$) is constructed by relating the exact eigenvector ($\mathbf{c}^k$) to the approximate vector ($\mathbf{x}^k$) through a correction vector ($\boldsymbol{\delta}^k$),[12, 27]

$$\mathbf{c}^k = \mathbf{x}^k + \boldsymbol{\delta}^k, \tag{5}$$

where $\boldsymbol{\delta}^k$ satisfies

$$(\mathbf{H} - \lambda^k)\boldsymbol{\delta}^k = -(\mathbf{H} - \lambda^k)\mathbf{x}^k = -\mathbf{r}^k \tag{6}$$

and $\mathbf{r}^k$ is the residual vector. If the current iteration's approximation (ρ^k) to the kth root is substituted for λ^k, then $\mathbf{H} - \lambda^k$ is no longer singular, and eq. (6) constitutes a traditional, shifted inverse iteration approach[11] to computing $\boldsymbol{\delta}^k$ from an arbitrary starting guess. In essence, as shown by spectral decomposition of **H**, those components of $\mathbf{r}^k$ involving eigenvectors with eigenvalues in close proximity to ρ^k are selectively magnified in the update vector $\boldsymbol{\delta}^k$. From another viewpoint, eq. (6) with the ρ^k replacement amounts to a Newton–Raphson update for finding eigenvectors as stationary points of the Rayleigh quotient

$$\rho = \frac{\boldsymbol{x}^T H \boldsymbol{x}}{\boldsymbol{x}^T \boldsymbol{x}} \tag{7}$$

provided that residual-vector terms are neglected in the Hessian of ρ. Unfortunately, eq. (6) involves solving a system of linear equations with the same dimension as the CI eigenvalue problem *for each iteration and for each root sought*. This difficulty is circumvented by using the sparsity of the CI Hamiltonian matrix and approximating **H**, on the left side of eq. (6), by some approximate, invertible Hamiltonian $\mathbf{H}_0$. In brief, the correction vector is derived by applying the preconditioner matrix $(\mathbf{H}_0 - \rho^k)^{-1}$ onto the current residual vector, i.e., $\boldsymbol{\delta}^k = -(\mathbf{H}_0 - \rho^k)^{-1}\boldsymbol{r}^k$. Diagonal dominance of the CI matrix leads in the simplest approximation to setting $\mathbf{H}_0$ to the diagonal part, **D**, of **H**:[12, 27]

$$\boldsymbol{\delta}^k = -(\mathbf{D} - \rho^k)^{-1}(\mathbf{H} - \rho^k)\mathbf{x}^k = -(\mathbf{D} - \rho^k)^{-1}\mathbf{r}^k \tag{8}$$

The resulting $\boldsymbol{\delta}^k$ vector is orthonormalized against the list of previous expansion vectors and appended to the expansion space. The iterative process is considered converged when a threshold value of $10^{-\eta}$ is reached, where η is typically 6–10 for the energy or 3–6 for the norm of the residual vector. Because the residual vector is computed each iteration according to eq. (9), it is necessary to store all previous **b** and $\boldsymbol{\sigma}$ vectors.

$$\mathbf{r}^k = \sum_{i=1}^{L} \alpha_i^k (\boldsymbol{\sigma}_i - \rho^k \mathbf{b}_i) \tag{9}$$

In the computation of excited states, the original Davidson method solves for higher roots in a sequential fashion.[12] To reduce the amount of I/O operations, Liu suggested solving for all roots of interest simultaneously.[27] This approach has the advantage that the expansion vectors for any given root provide additional variational flexibility and energy minimization within the subspace of the **G** matrix for all roots sought. This modification of the original Davidson method is known by several names, including the Block Davidson, Davidson–Liu, and simultaneous expansion method.[27, 28] A general outline for this algorithm is given in Appendix A.

SUBSPACE COLLAPSE

An increase in disk I/O may substantially degrade the effectiveness of the diagonalization algorithm. Typically, I/O is reduced by restarting the Davidson procedure with the latest eigenvector guess when the number of expansion vectors becomes too large, which is equivalent to a collapse of the $\mathbf{b}_i$ subspace to one dimension. However, this truncation can hinder convergence due to loss of variational flexibility. In 1990, van Lenthe and Pulay[29] reported that the convergence of the diagonalization procedure is nearly conserved if the expansion space is collapsed to two vectors *each iteration* rather than to one vector every several iterations. This remarkable behavior can be justified[29] from conjugate gradient theory, assuming the collapse is accomplished by taking the CI eigenvectors given by eq. (2) from the last two iterations and orthonormalizing them. Murray, Racine, and Davidson[28] generalized this approach for excited states and also proposed a complementary scheme entailing periodic least-squares extrapolation to minimize the norm of the residual vector under the constraint of a fixed eigenvalue and eigenvector norm. The alternative expansion coefficients ($\boldsymbol{\alpha}^k$) resulting from this extrapolation are obtained as the eigenvectors of the root-dependent residual

overlap matrix:

$$M_{ij} = \left[(\mathbf{H} - \lambda^k)\mathbf{b}_i\right]^T (\mathbf{H} - \lambda^k)\mathbf{b}_j \qquad (10)$$

Murray et al.[28] reported a 20% increase in efficiency for obtaining 10 roots of a CI matrix when the extrapolation was invoked every third iteration after the energy correction reached a threshold of 5.0×10^{-3} E_h. A periodic collapse to two vectors rather than one vector per root when the expansion space reached 100 vectors yielded an additional 20% savings.

CORRECTION VECTOR

Several investigators have explored modifications of the correction/update vector in the diagonalization procedure. In 1990, Olsen[15] constructed a correction vector that ensures the introduction of new character into the expansion space near convergence. Analysis of the Davidson method reveals that as $\mathbf{H}_0 \rightarrow \mathbf{H}$, $\boldsymbol{\delta}^k \rightarrow -\boldsymbol{x}^k$, raising concerns of linear dependence in the expansion vectors.[15, 23, 25, 30–32] The Olsen correction scheme in the $\mathbf{H}_0 \rightarrow \mathbf{H}$ limit gives

$$\boldsymbol{\delta}^k = -\mathbf{x}^k + \frac{(\mathbf{H}_0 - \rho^k)^{-1}\mathbf{x}^k}{(\mathbf{x}^k)^T(\mathbf{H}_0 - \rho^k)^{-1}\mathbf{x}^k}, \qquad (11)$$

in which $\boldsymbol{\delta}^k$ is explicitly orthogonal to and hence linearly independent of $\mathbf{x}^k$. The general form of Olsen correction vector is

$$\boldsymbol{\delta}^k = -\left(\mathbf{H}_0 - \rho^k\right)^{-1}\left(\mathbf{H} - \rho^k_{\text{new}}\right)\mathbf{x}^k, \qquad (12)$$

which differs from the Davidson vector [eq. (8)] only by the substitution of an updated eigenvalue estimate (ρ^k_{new}) containing first-order corrections.[15] Equation (12) can be derived by a simple perturbation analysis in which the change in the eigenvalue estimate is determined directly from the condition $\boldsymbol{\delta}^k \cdot \mathbf{x}^k = 0$. The Olsen update can also be derived by first invoking the Newton–Raphson method in search of stationary points of a Lagrangian function related to the Rayleigh quotient and then substituting the preconditioner as an estimate of $(\mathbf{H} - \rho^k)^{-1}$.[33]

In 1990, the Olsen correction vector scheme was employed in a single-vector iterative diagonalization algorithm to perform the first iterations of a one billion determinant full CI computation.[15] Although the Olsen single-vector method is necessary if external storage is limited, its convergence characteristics are not always ideal. Mitrushenkov[26] emphasized that Olsen's iterator does not determine expansion coefficients through a minimization procedure. Mitrushenkov proposed a modification to Olsen's scheme, which avoids this difficulty by utilizing a nonorthogonal two-vector version of the algorithm, whereby $\boldsymbol{\delta}^k$ and $\boldsymbol{x}^k$ are optimally mixed by diagonalization of a corresponding 2×2 subspace matrix $\mathbf{G}$.[26] Bofill and Anglada[33] have derived the Olsen correction vector from other considerations and combined it with a two-vector subspace collapse in a Lanczos-type method. Their approach was found to converge more rapidly than the method of van Lenthe and Pulay,[29] where the Davidson correction vector was employed. The application of this method to excited states revealed that the optimization of any root can be achieved without solving for the lower roots of the same spin and spatial symmetry.[33] Bendazzoli and Evangelisti[34] utilized a single-vector variant of the Davidson and Olsen methods wherein the correction vector is derived from

$$\boldsymbol{\delta}^k = -\left[\mathbf{H} - \rho^k - \mathbf{r}^k\left(\mathbf{x}^k\right)^T - \mathbf{x}^k\left(\mathbf{r}^k\right)^T + \beta\mathbf{x}^k\left(\mathbf{x}^k\right)^T\right]^{-1}\mathbf{r}^k, \qquad (13)$$

where β is an arbitrary, nonzero parameter whose choice does not noticeably affect convergence of the algorithm. If β is chosen to be ρ^k, eq. (13) is equivalent to the Jacobi–Davidson update scheme. Equation (13) is an approximation to a formal solution[35] of eq. (6) in the orthogonal complement of the null space of $\mathbf{H} - \lambda^k$. Again, the update vector (ignoring the β term) can be viewed as a Newton–Raphson iteration for stationary points of the Rayleigh quotient, this time the residual terms being included in the Hessian of ρ, unlike the Davidson scheme. In practice, only diagonal elements have been taken in forming the preconditioner of eq. (13). This approach has been employed in the computation of full CI benchmark energies that involve up to ten billion determinants to μE_h accuracy.[16–18, 36–39]

GENERALIZED DAVIDSON METHOD

The performance of the diagonalization algorithm can be improved by lifting the assumption of diagonal dominance in the preconditioner through the explicit inclusion of coupling between the most important determinants. This generalized Davidson preconditioner improves convergence considerably when there are near degeneracies in the CI spectrum.[25, 40] A partitioned $\mathbf{H}$ matrix in eq. (6) may be selected by invoking three model spaces according to the value of the diagonal Hamiltonian matrix elements: reference space 0 containing the most important determinants and full internal coupling among them; a larger, less important interacting space 1 possessing external coupling to the reference space but no internal coupling; and the remaining, least important space 2 exhibiting neither internal

nor external coupling. With λ^k approximated by ρ^k, eq. (6) is then reduced to

$$\begin{bmatrix} \mathbf{H}_{00}-\rho^k & \mathbf{H}_{01} & 0 \\ \mathbf{H}_{10} & \mathbf{D}_{11}-\rho^k & 0 \\ 0 & 0 & \mathbf{D}_{22}-\rho^k \end{bmatrix} \begin{bmatrix} \boldsymbol{\delta}_0^k \\ \boldsymbol{\delta}_1^k \\ \boldsymbol{\delta}_2^k \end{bmatrix} = - \begin{bmatrix} \mathbf{r}_0^k \\ \mathbf{r}_1^k \\ \mathbf{r}_2^k \end{bmatrix}. \tag{14}$$

In the simplest case, space 1 is null, and the preconditioner is obtained by trivial inversion of the diagonal matrix $\mathbf{D}_{22}-\rho^k$, and the use of standard direct[10] or iterative methods[10, 41] for solving the $\mathbf{H}_{00}-\rho^k$ linear system for $\boldsymbol{\delta}_0^k$. Otherwise, the exact solution for $\boldsymbol{\delta}_0^k$ and $\boldsymbol{\delta}_1^k$ can be represented as

$$\begin{aligned} -\boldsymbol{\delta}_0^k &= \left(\mathbf{H}_{00}-\rho^k\right)^{-1}\left[(\mathbf{I}-\boldsymbol{\Delta}_0\boldsymbol{\Delta}_1)^{-1}\mathbf{r}_0^k \right. \\ &\qquad \left. -\boldsymbol{\Delta}_0(\mathbf{I}-\boldsymbol{\Delta}_1\boldsymbol{\Delta}_0)^{-1}\mathbf{r}_1^k\right], \\ -\boldsymbol{\delta}_1^k &= \left(\mathbf{D}_{11}-\rho^k\right)^{-1}\left[(\mathbf{I}-\boldsymbol{\Delta}_1\boldsymbol{\Delta}_0)^{-1}\mathbf{r}_1^k \right. \\ &\qquad \left. -\boldsymbol{\Delta}_1(\mathbf{I}-\boldsymbol{\Delta}_0\boldsymbol{\Delta}_1)^{-1}\mathbf{r}_0^k\right], \end{aligned} \tag{15}$$

where

$$\begin{aligned} \boldsymbol{\Delta}_0 &= \mathbf{H}_{01}\left(\mathbf{D}_{11}-\rho^k\right)^{-1}, \\ \boldsymbol{\Delta}_1 &= \mathbf{H}_{10}\left(\mathbf{H}_{00}-\rho^k\right)^{-1}. \end{aligned} \tag{16}$$

For limited coupling, $\boldsymbol{\Delta}_0$ and $\boldsymbol{\Delta}_1$ are small; thus, they may be neglected, or employed in the well-known matrix expansion

$$(\mathbf{I}-\mathbf{R})^{-1} = \mathbf{I}+\mathbf{R}+\mathbf{R}^2+\cdots \tag{17}$$

where

$$\mathbf{R} = \boldsymbol{\Delta}_0\boldsymbol{\Delta}_1 \qquad \text{or} \qquad \boldsymbol{\Delta}_1\boldsymbol{\Delta}_0. \tag{18}$$

While all terms in low-order expansions of eq. (15) are computationally accessible, the extra effort exerted to obtain the best preconditioner is rarely warranted.

APPROXIMATE DIAGONAL ELEMENTS

For large-scale computations it is advantageous to approximate the diagonal Hamiltonian matrix elements in the preconditioner to maintain the spin symmetry of the expansion basis, and to allow efficient on-the-fly evaluation. In a Slater determinant basis it is possible to introduce spin contamination in the expansion vectors during the preconditioning step [eq. (8) and eq. (27)], a bothersome feature that may plague the eigenvector estimate until full convergence is reached. Spin-adapted configuration state functions (CSFs) are formed from a set of determinants having the same *spatial* orbital occupation pattern but with different spin couplings. The determinants with identical spatial orbital occupancies will couple to form CSFs with various values of the spin quantum number S.[42, 43] Let us call this set of determinants a spin-coupling set. For example, the spin-coupling set with spatial orbitals p, q, r, and s singly occupied outside a closed shell is $|pq\bar{r}\bar{s}\rangle$, $|p\bar{q}r\bar{s}\rangle$, $|p\bar{q}\bar{r}s\rangle$, $|\bar{p}qr\bar{s}\rangle$, $|\bar{p}q\bar{r}s\rangle$, $|\bar{p}\bar{q}rs\rangle$. All determinants in a spin-coupling set must have certain fixed relationships for spin eigenfunctions to result. For $M_s = 0$ states these relationships include the time-reversal symmetry constraint[9]

$$C(I_\alpha, I_\beta) = (-1)^S C(I_\beta, I_\alpha), \tag{19}$$

where I_α and I_β denote the occupation lists for the α and β strings.[3] In our previous example, eq. (19) requires that the CI coefficients for determinants (1,6), (2,5), and (3,4) of the list differ internally only by the factor $(-1)^S$.

If the spin eigenfunction relationships are satisfied by the trial eigenvector, then both the $\boldsymbol{\sigma}$ and residual vectors will also have the proper relationships because the nonrelativistic CI Hamiltonian does not contain spin-dependent operators. Thus, if the correction vector is to have the proper relationships among determinantal coefficients belonging to a spin-coupling set, then the denominator for *every* determinant in a spin-coupling set must be *identical*, otherwise the Davidson subspace vectors will not be spin eigenfunctions. It should be noted that for the generalized Davidson method each element in a spin-coupling set must be included in the same model space, or spin contamination will result in the expansion vectors.

An average diagonal energy for a spin-coupling set can be derived as follows. First, let us consider the diagonal energy for a given single determinant. For a given determinant, (n_i^α, n_i^β) will represent a binary digit (1 or 0) signifying whether an electron with (α, β) spin is present in orbital i. The diagonal energy in terms of the usual one-electron Hamiltonian (h_{ii}), Coulomb (J_{ij}), and exchange (K_{ij}) integrals is then

$$\begin{aligned} H_{II} = &\sum_i \left(n_i^\alpha + n_i^\beta\right)h_{ii} + \frac{1}{2}\sum_{ij} n_i^\alpha n_j^\alpha J_{ij} \\ &+ \frac{1}{2}\sum_{ij} n_i^\beta n_j^\beta J_{ij} + \sum_{ij} n_i^\alpha n_j^\beta J_{ij} \\ &- \frac{1}{2}\sum_{ij} n_i^\alpha n_j^\alpha K_{ij} - \frac{1}{2}\sum_{ij} n_i^\beta n_j^\beta K_{ij}, \end{aligned} \tag{20}$$

where the summations run over all the (ROHF) spatial orbitals. The energy of each determinant in a spin-coupling set differs through the exchange integrals, as in our previous example, where the open-shell exchange contribution to the energy of the six determinants is $(-K_{pq}-K_{rs}, -K_{pr}-K_{qs}, -K_{ps}-$

$K_{qr}, -K_{ps} - K_{qr}, -K_{pr} - K_{qs}, -K_{pq} - K_{rs}$), in order. The most accurate approximation is to employ an average of all the exchange integrals in a spin-coupling set. In our example, this approach ascribes an exchange contribution of $-(\frac{1}{3})(K_{pq} + K_{rs} + K_{pr} + K_{qs} + K_{ps} + K_{qr})$ to the diagonal element of the preconditioner for all six determinants. Knowles and Handy[8] advocate the use of modified one-electron integrals, which fold in part of the two-electron exchange integral contribution, and replace the exchange integrals by the maximum exchange integral for the molecule. However, we found this approximation to nearly double the number of iterations in the Davidson procedure, and because of the numerous integral look-ups it does not allow efficient on-the-fly evaluation. A more viable alternative is the approximation

$$H_{II} = \frac{1}{2}\sum_i (n_i^\alpha + n_i^\beta)(\epsilon_i + h_{ii}), \tag{21}$$

which deviates from the correct expression in that the two-electron terms $\langle rs||rs\rangle$ for excited spin orbitals r and s of the Hartree–Fock virtual space are reckoned as $\langle rh||rh\rangle$, where spin orbital h is a corresponding hole in the Hartree–Fock occupied space.

The concept of approximating or averaging diagonal matrix elements to enhance algorithmic performance or preserve spin properties has existed in the literature for some time.[14, 44] Recently, Evangelisti et al.[16, 38] have suggested several schemes that enable efficient on-the-fly evaluation of average diagonal Hamiltonian elements, thus eliminating the need to store the entire diagonal vector on disk or in core memory. The first approach[38] is simply to use the sum of orbital energies (ϵ_i) of the occupied orbitals for each determinant, a relatively crude estimate, which, unlike eq. (21), does not attempt to correct for overcounting of two-electron interactions. A more accurate approach[16] is to evaluate a diagonal element as

$$D(I_\alpha, I_\beta) = E_{\mathrm{HF}} + F(I_\alpha) + F(I_\beta), \tag{22}$$

where $F(I)$ for occupation string I contains an orbital-energy correction to the Hartree–Fock reference energy (E_{HF}) arising from the differences in occupations between the Hartree–Fock and excited determinants. Specifically,

$$F(I) = \sum_{a\in I}^{\mathrm{vir}} \epsilon_a - \sum_{a\notin I}^{\mathrm{occ}} \epsilon_i, \tag{23}$$

where occ and vir refer to the occupied and virtual spaces of the reference configuration. The dimension of the $F(I)$ array is much smaller than the size of the CI problem, for example, in a full CI computation the length of $F(I)$ is approximately the square root of the number of determinants. Therefore, the $F(I)$ array can be precomputed and stored in core memory. Both Evangelisti schemes for approximating diagonal elements, as well as eq. (21), are computationally efficient, and their convergence effects on the diagonalization procedure will be discussed shortly.

CURRENT STUDY

In this article we calibrate the performance of several modifications to the Davidson algorithm against the traditional Davidson method for an assortment of atomic and molecular computations. We consider how the convergence characteristics are affected by a generalized Davidson method, a periodic n-vector subspace collapse, a least-squares extrapolation technique, several one- and two-vector approaches, alternative correction vectors, and approximate diagonal Hamiltonian matrix elements in the preconditioner. Although compression schemes[39, 45–47] are useful for reducing I/O requirements of large-scale computations, we exclude them from the present study. Indeed, to keep the scope of our investigation manageable, the algorithms tested here are not exhaustive of the plethora of large matrix methods in use by the scientific community at large. Rather, our work is focused on establishing a diagonalization scheme for electronic structure problems that reduces disk storage and I/O requirements, but maintains the convergence characteristics of the full subspace scheme. Such an approach would permit the computation of large-scale CI wave functions with the convergence criteria necessary for the calculation of general properties over a wider range of chemical systems.

Computational Details

In this study computations on the chemical systems in Table I were used to calibrate the performance of the diagonalization methods. The computations utilized the correlation consistent cc-pVDZ and cc-pVQZ basis sets of Dunning and coworkers,[48] as well as DZP and TZ2P sets. The DZP basis sets were comprised of the standard Huzinaga–Dunning[49] double-ζ sp set augmented with a d-type manifold [α_d(C) = 0.75, α_d(O) = 0.85, α_d(S) = 0.70], while the TZ2P basis consisted of the standard Huzinaga–Dunning[50] triple-ζ sp set augmented with two polarization manifolds [α_p(H) = 1.5, 0.375; α_d(O) = 1.7, 0.4250;

TABLE I.

Iterations Required by Various Diagonalization Schemes for (10^{-10} E_h) Convergence of Selected CI Wave Functions for Model Species.[a]

	(Collapse Size, Subspace Dimension)[b]					
(Precon, Update)	(∞, ∞)	(2,4)	(2,3)	(1,4)	(1,3)	(1,2)
$\tilde{X}\ ^1A_1$ H_2O (cc-pVDZ FCI)	19 604 169 dets (C_{2v})					
(DVD, DVD)	12	12 (13)	12 (13)	14 (14)	15 (14)	22
(DVD, OLS)	12	12 (13)	12 (13)	14 (14)	16 (14)	22
(GDVD, DVD)	10	10 (11)	10 (10)	11 (11)	11 (11)	15
(GDVD, OLS)	10	10 (11)	10 (10)	11 (11)	11 (11)	15
(GDVD, OLS)	OLS(1,1) = 25 (10 for μE_h conv)[c] MIT(1,2) = 21 (7 for μE_h conv)[c]					
$\tilde{X}\ ^1A_1$ O_3 (TZ2P CISD)	109 033 dets (C_{2v})					
(DVD, DVD)	13	13 (14)	14 (14)	15 (15)	17 (14)	26
(DVD, OLS)	13	13 (14)	13 (14)	15 (15)	16 (14)	26
(GDVD, DVD)	11	12 (12)	12 (13)	13 (12)	13 (13)	22
(GDVD, OLS)	11	12 (12)	11 (12)	12 (12)	12 (12)	17
(GDVD, OLS)	OLS(1,1) = 21 (12 for μE_h conv) MIT(1,2) = 20 (11 for μE_h conv)					
$X\ ^1\Sigma_g^+$ N_2 (cc-pVDZ CISDTQ)	969 718 dets (D_{2h})					
(DVD, DVD)	15	15 (16)	15 (16)	17 (16)	18 (16)	28
(DVD, OLS)	14	15 (16)	15 (16)	17 (16)	18 (16)	28
(GDVD, DVD)	12	13 (12)	12 (12)	13 (13)	13 (12)	17
(GDVD, OLS)	12	13 (12)	12 (12)	13 (13)	13 (12)	17
(GDVD, OLS)	OLS(1,1) = 17 (9 for μE_h conv) MIT(1,2) = 17 (9 for μE_h conv)					
$X\ ^1\Sigma_g^+$ C_2 (DZP CISDTQ)	582 455 dets (D_{2h})					
(DVD, DVD)	18	18 (22)	19 (26)	31 (21)	33 (26)	60
(DVD, OLS)	18	18 (23)	19 (26)	31 (21)	36 (26)	60
(GDVD, DVD)	14	14 (15)	14 (16)	18 (16)	20 (17)	28
(GDVD, OLS)	14	14 (15)	14 (16)	16 (16)	19 (17)	28
(GDVD, OLS)	OLS(1,1) = DNC[d] (DNC[d] for μE_h conv)[c] MIT(1,2) = DNC[d] (10 for μE_h conv)[c]					
$2\ ^1\Sigma_g^+$ C_2 (DZP CISDTQPH)	16 786 215 dets (D_{2h})					
(DVD, DVD)	20	21 (25)	21 (25)	28 (26)	29 (26)	86 (32)
(DVD, OLS)	20	21 (25)	21 (25)	28 (26)	30 (26)	86 (32)
(GDVD, DVD)	16	16 (19)	16 (20)	18 (18)	20 (18)	36 (22)
(GDVD, OLS)	16	16 (18)	16 (18)	19 (18)	19 (18)	35 (20)
$\tilde{X}\ ^1A_{1g}$ C_2H_4 (TZ2P CISD)	40 021 dets (D_{2h})					
(DVD, DVD)	11	11 (11)	11 (11)	11 (11)	11 (11)	16
(DVD, OLS)	11	11 (11)	11 (11)	11 (11)	11 (11)	15
(GDVD, DVD)	9	9 (9)	10 (10)	10 (10)	9 (9)	16
(GDVD, OLS)	9	10 (10)	10 (10)	10 (10)	9 (9)	12
(GDVD, OLS)	OLS(1,1) = 14 (8 for μE_h conv) MIT(1,2) = 14 (8 for μE_h conv)					

TABLE I.
(Continued)

		(Collapse Size, Subspace Dimension)[b]				
(Precon, Update)	(∞, ∞)	(2,4)	(2,3)	(1,4)	(1,3)	(1,2)
$\tilde{X}\ ^1A_1$ SO_2 (DZP CISD)	41 738 dets (C_{2v})					
(DVD, DVD)	13	13 (14)	13 (14)	15 (13)	15 (14)	25
(DVD, OLS)	13	13 (14)	13 (14)	15 (13)	15 (14)	25
(GDVD, DVD)	11	11 (11)	11 (12)	12 (12)	12 (12)	21
(GDVD, OLS)	11	11 (11)	11 (11)	11 (11)	11 (11)	16
(GDVD, OLS)	OLS(1,1) = 20 (11 for μE_h conv) MIT(1,2) = 18 (8 for μE_h conv)					
$\tilde{X}\ ^1\Sigma^+$ HCN (cc-pVDZ CISDTQ)	4 601 819 dets (C_{2v})					
(DVD, DVD)	15	15 (15)	15 (16)	16 (16)	18 (16)	25
(DVD, OLS)	15	15 (15)	15 (16)	16 (16)	19 (16)	25
(GDVD, DVD)	13	13 (13)	13 (13)	13 (13)	14 (13)	18
(GDVD, OLS)	13	13 (13)	13 (13)	13 (13)	14 (13)	19
(GDVD, OLS)	OLS(1,1) = 87 (13 for μE_h conv) MIT(1,2) = 72 (10 for μE_h conv)					
$X\ ^1\Sigma^+$ SiO (cc-pVDZ CISDTQ)	1 938 774 dets (C_{2v})					
(DVD, DVD)	16	16 (17)	16 (18)	19 (18)	23 (19)	34
(DVD, OLS)	16	16 (17)	16 (18)	19 (18)	23 (19)	34
(GDVD, DVD)	14	14 (15)	14 (15)	16 (16)	19 (16)	27
(GDVD, OLS)	14	14 (15)	15 (15)	16 (16)	19 (16)	27
(GDVD, OLS)	OLS(1,1) $\geq$ 100 (33 for μE_h conv) MIT(1,2) $\geq$ 100 (12 for μE_h conv)					
$\tilde{X}\ ^1\Sigma_g^+$ C_2H_2 (cc-pVDZ CISDTQ)	4 681 508 dets (D_{2h})					
(DVD, DVD)	13	13 (14)	13 (14)	14 (15)	17 (14)	22
(DVD, OLS)	13	13 (14)	13 (14)	14 (15)	17 (14)	22
(GDVD, DVD)	11	11 (12)	11 (12)	12 (12)	13 (12)	17
(GDVD, OLS)	11	11 (12)	11 (12)	12 (12)	13 (12)	17
(GDVD, OLS)	OLS(1,1) = 38 (12 for μE_h conv) MIT(1,2) = 28 (10 for μE_h conv)					
$\tilde{X}\ ^1A_1$ NH_3 (cc-pVDZ CISDTQ)	2 333 495 dets (C_s)					
(DVD, DVD)	13	13 (14)	13 (14)	16 (14)	18 (15)	27
(DVD, OLS)	13	13 (14)	13 (14)	16 (14)	18 (16)	26
(GDVD, DVD)	11	11 (12)	11 (12)	12 (12)	14 (13)	20
(GDVD, OLS)	11	11 (11)	11 (11)	12 (12)	13 (12)	17
(GDVD, OLS)	OLS(1,1) = 23 (11 for μE_h conv) MIT(1,2) = 20 (8 for μE_h conv)					

α_d(C) = 1.5, 0.375]. The C_2 computations employed Cartesian polarization sets while the remaining test cases utilized pure spherical harmonics. The core (canonical Hartree–Fock) orbitals were frozen for all computations, and for C_2 the corresponding $1s^*$ virtual orbitals were deleted. Configuration interaction wave functions through double (CISD), quadruple (CISDTQ), and hextuple (CISDTQPH) substitutions were included in the calibration. The computations were carried out at the following structures: [0.95885 Å, 104.34°] (H_2O), [1.2808 Å, 116.71°] (O_3), 1.1996 Å (N_2), 1.2695 Å (C_2), [1.076 Å (r_{CH}), 1.330 Å (r_{CC}), 116.6° (θ_{HCH})] (C_2H_4), [1.4303 Å, 119.28°] (SO_2), [1.064 Å (r_{CH}), 1.156 Å (r_{CN})] (HCN),

TABLE I.
(Continued)

	(Collapse Size, Subspace Dimension)[b]					
(Precon, Update)	(∞, ∞)	(2,4)	(2,3)	(1,4)	(1,3)	(1,2)
$X\ ^1\Sigma_g^+$ F_2 (cc-pVQZ CISD)	90 926 dets (D_{2h})					
(DVD, DVD)	12	12 (13)	12 (13)	13 (13)	13 (13)	20
(DVD, OLS)	10	12 (13)	12 (13)	13 (13)	13 (13)	20
(GDVD, DVD)	10	9 (9)	10 (10)	10 (10)	9 (9)	16
(GDVD, OLS)	10	9 (9)	9 (9)	10 (10)	9 (9)	11
(GDVD, OLS)	OLS(1,1) = 15 (9 for μE_h conv) MIT(1,2) = 13 (7 for μE_h conv)					

[a] Both the initial guess vector and the generalized Davidson preconditioner were obtained from model spaces ($\mathbf{H}_{00}$) of 400 and 1000 determinants for CI spaces below and above one million determinants, respectively. No intermediate interacting space 1 in eq. (14) was used.
[b] Entries in parentheses involve least-squares extrapolation every third iteration.
[c] To achieve facile convergence, the model space was increased to 750 determinants.
[d] Did not converge.

1.50974 Å (SiO), [1.06126 Å (r_{CH}), 1.2041 Å (r_{CC})] (C_2H_2), [1.0138 Å (r_{NH}), 106.13° (θ_{HNH})] (NH_3), and 1.41193 Å (F_2). All computations were executed with the PSI3 package[51] as linked to the determinant-based CI program DETCI.[3]

Results and Discussion

The data obtained in this study on the performance of various diagonalization schemes and myriad variants thereof are summarized in Tables I–III. The chemical systems chosen for investigation exhibit varying degrees of multireference character, ranging from the classic singly bonded, 10-electron systems H_2O and NH_3, through the closed-shell, multiply bonded C_2H_4, C_2H_2, N_2, and HCN molecules, to species such as O_3 recognized for strong diradical character, and finally to the extreme case of the two lowest $^1\Sigma_g^+$ states of C_2. As manifested in our earlier full CI and very high-order MPn studies,[52, 53] the carbon dimer exhibits an intricate electronic structure due to a near degeneracy of the fifth σ orbital and first π orbital.

Specification of the various diagonalization schemes starts with the choice of the CI correction vector, given either by the Davidson [DVD, eq. (8)] or Olsen [OLS, eq. (12)] update prescription. For the next option, the generalized Davidson (GDVD) preconditioner (precon) includes off-diagonal coupling terms in eq. (14), while the original Davidson (DVD) approach includes only diagonal elements.

TABLE II.
Performance of Zeroth-Order Blocksize in the Generalized Davidson Preconditioner.[a]

$\mathbf{H}_{00}$ Blocksize/Dets	$\tilde{X}\ ^1A_1$ H_2O 19,604,169	$2\ ^1\Sigma_g^+$ C_2 16,786,215	$\tilde{X}\ ^1\Sigma^+$ HCN 4,601,819	$X\ ^1\Sigma^+$ SiO 1,938,774	$\tilde{X}\ ^1\Sigma_g^+$ C_2H_2 4,681,508	$\tilde{X}\ ^1A_1$ NH_3 2,333,495
1	12	21	15	16	13	13
400	10	18	12	14	11	11
1000	10	16	12	13	11	11
2000	10	15	11	13	11	10
3000	9	15	11	13	10	9
4000	9	15	11	12	10	9
10,000	9	15	11	12	10	9

[a] Performance is quantified in the table by the number of iterations necessary for 10^{-10} E_h convergence. All computations involved the same CI spaces specified in Table I and utilized a (precon, update) = (GDVD, DVD) approach with a (2,3) collapse scheme.

TABLE III.
Performance of Diagonal Element Approximations in the Preconditioner.[a]

Method[b]	$\tilde{X}\,^1A_1$ H_2O	$X\,^1\Sigma_g^+$ C_2	$\tilde{X}\,^1\Sigma^+$ HCN	$X\,^1\Sigma^+$ SiO	$\tilde{X}\,^1\Sigma_g^+$ C_2H_2	$\tilde{X}\,^1A_1$ NH_3
A	11	15	12	14	12	11
B	10	15	12	14	11	11
C	11	16	12	14	12	12
D	12	20	14	19	14	12
E	18	44	20	23	19	17

[a] Performance is quantified in the table by the number of iterations required for 10^{-10} E_h convergence. All computations involved the same CI spaces specified in Table I and utilized a (GDVD, DVD) approach with a (2,3) collapse scheme. An $\mathbf{H}_{00}$ block size of 1000 was used throughout. The various approximations were only applied to diagonal elements of the Hamiltonian outside the $\mathbf{H}_{00}$ block.
[b] A = exact diagonal energies; B = diagonal energies with averaged exchange integrals within each spin-coupling set; C = Evangelisti approximation, eq. (22); D = corrected orbital energy approximation, eq. (21); E = simple orbital energy sum approximation.

Combining (precon, update) choices yields four basic methods for testing in Table I. Variations of these methods are denoted by the notation (x, y), which signifies collapsing the vector subspace to x vectors per root with a maximum of y vectors per root in the subspace prior to collapse. For comparison purposes, secondary diagonalization runs for each algorithm variant were performed by invoking least-squares extrapolation via eq. (10) of the residual vectors every third iteration after the energy had converged to 10^{-4} E_h. Finally, the specialized Olsen [OLS(1,1)][15] and Mitrushenkov [MIT(2,1)][26] methods designed for exceptionally large CI problems were added to the set of algorithms subjected to performance testing. For all computations reported in Table I, the initial guess vector was selected from the diagonalization of a small subblock of the Hamiltonian matrix, chosen according to the lowest diagonal elements, with dimension ranging from 400 to 1000 determinants. In addition, the diagonal elements in the preconditioning matrix were approximated, by averaging the exchange integrals over entire spin-coupling sets, thereby maintaining spin symmetry of the CI vectors at each iteration. In our analysis of the performance data in Table I, we focus on percent variations in the number of iterations required to achieve 10^{-10} E_h convergence, although this simple criterion is certainly not the only factor to consider when comparing different methods. For example, additional I/O operations incur a significant amount of overhead and, therefore, must also be considered in the selection of the most efficient methods, an aspect addressed separately below. The relatively stringent 10^{-10} E_h primary threshold is representative of the convergence levels required to tightly optimize geometric structures, to compute force fields by finite difference schemes, or to accurately evaluate various molecular properties; it also more clearly exposes differences in algorithmic capabilities. On the other hand, a 10^{-6} E_h secondary threshold is used here to assess performance in circumstances only requiring chemically significant comparisons of single-point energies.

The first observation in Table I is that for all molecules the traditional (precon, update) = (DVD, DVD) method with no subspace restrictions (∞, ∞) converges the energy to 10^{-10} E_h in 12–20 iterations. By adding coupling in the preconditioner, the (GDVD, DVD) method decreases the number of required iterations by a substantial $17 \pm 5\%$, the largest improvement occurring, as expected, for the multireference C_2 system. Without subspace restriction, the choice of the DVD vs. OLS update vector has virtually no effect on performance, as in only two cases do the [(DVD, DVD), (DVD, OLS)] results differ and in no cases do the [(GDVD, DVD), (GDVD, OLS)] entries vary. The superiority of the GDVD preconditioner waxes with truncation of the vector space, particularly if the OLS update is employed. In the extreme (1,2) truncation, the GDVD reduction in the number of iterations, for the 10 cases excluding C_2, grows to (22%, 31%) for the (DVD, OLS) updates. More strikingly, for $X\,^1\Sigma_g^+$ C_2, the iterations required with (1,2) truncation are reduced from 60 to 28 by the GDVD preconditioner, regardless of the update vector, and the number of iterations saved in the $2\,^1\Sigma_g^+$ C_2 case is even larger.

The data in Table I for the (x, y) truncation series provide both unequivocal support for the efficacy of subspace collapse and clear demonstration of pitfalls to be avoided in this process. For both the (precon, update) = (DVD, DVD) and (GDVD, DVD) algorithms, the (2,4) and (2,3) collapse schemes increase the number of iterations over the (∞, ∞)

benchmark by less than 1 and 2%, respectively. This remarkable performance shows that the full convergence characteristics of the unrestricted Davidson approach can indeed be maintained with subspaces no larger than three expansion vectors. Truncation beyond the (2,3) level does come with a cost, however. For the 10 cases excluding C_2, the (1,4), (1,3), and (1,2) collapses cause 13, 22, and 83% increases over (∞, ∞) within the (DVD, DVD) method. Even with preconditioner coupling in the (GDVD, DVD) approach, these same iteration increases due to truncation remain at 9, 12, and 69%, in order. Moreover, the C_2 cases reveal the possible deleterious effects of all procedures that collapse to only a single vector; for example, the set of [(1,4), (1,3), (1,2)] truncations for X $^1\Sigma_g^+$ C_2 results in prodigious (72, 83, 330%) and (29, 43, 100%) increases for the (DVD, DVD) and (GDVD, DVD) methods, in order. From the data in Table I we may conclude that (2,3) truncation provides no noticeable deterioration in performance at all; (1,3) truncation may result in efficiency loses less than 25%, but only for well-behaved systems; and (1,2) truncation is expected to require at least 50% more iterations, with possible catastrophic failure for multireference systems. Finally, one observes that there are no systematic differences in collapse behavior between the DVD and OLS update procedures, except in the (1,2) extreme with the GDVD preconditioner. In Table I, (GDVD, OLS) provides substantial savings of up to 31% over (GDVD, DVD) within the (1,2) scheme for O_3, C_2H_4, SO_2, NH_3, and F_2, but not for the remaining molecules. Thus, as noted above, for (1,2) collapse the OLS update appears to take greater advantage of the GDVD preconditioner in certain cases.

An assessment of least-squares extrapolation prior to collapse in the diagonalization procedure is provided by the entries in parentheses in Table I. If the limitations on the subspace are not extreme, as in the (2,4), (2,3), and (1,4) schemes, least-squares extrapolation generally has a neutral effect, with deterioration of perhaps an iteration for two-vector collapse but with comparable improvement for one-vector collapse. The multireference C_2 cases constitute vivid exceptions to this trend, especially if the GDVD preconditioner is not used. For example, for ground-state C_2 with the (DVD, DVD) method, least-squares extrapolation within the [(2,4), (2,3), (1,4)] schemes changes the number of required iterations from (18, 19, 31) to (22, 26, 21). As the limitations on the subspace becomes extreme, the extrapolation procedure gains real merit, once again more pronounced if only the DVD preconditioner is used. For the 12 cases presented in Table I, invoking extrapolation before collapse in the (DVD, DVD) (1,3) method reduces the number of iterations by an average of 11%. Finally, the 2 $^1\Sigma_g^+$ C_2 example demonstrates dramatic improvements afforded by least-squares extrapolation if the block Davidson algorithm is subjected to extreme (1,2) collapse for each root. To wit, with precon = DVD, extrapolation results in an immense 63% reduction in iterations, and even with precon = GDVD, the reduction is around 40%. This observation is consistent with the conclusions of Murray et al.[28] that if multiple roots are sought and the size of the vector space is limited, then least-squares extrapolation of residual vectors can substantially reduce the number of iterations in the block Davidson procedure.

The OLS(1,1) and MIT(2,1) diagonalization methods for very large CI problems are designed to sacrifice CPU time for reduced storage requirements, and the data reported in Table I reveal the extent of the compromise. For the ground states of C_2, HCN, and SiO, both methods prove problematic. They are effectively incapable of achieving tight (10^{-10} E_h) convergence for these molecules, minimally requiring over 70 iterations, or in the C_2 multireference system never reaching convergence at all. However, for the eight other examples in Table I, both diagonalization methods are viable for tight convergence. For OLS(1,1), the percentage increase in the number of iterations compared to the (GDVD, OLS) (∞, ∞) standard ranges from 50 to 250%, with an average of about 100%, among the eight well-behaved molecules. In the same comparison, the MIT(1,2) method reduces the mean increase to 74%. With respect to the (GDVD, OLS) (1,2) method of collapse, OLS(1,1) and MIT(1,2) are more competitive, but still require on average 40 and 23% more iterations, respectively. The OLS(1,1) and MIT(1,2) methods are most efficacious when only μE_h convergence is sought. For nine of the molecules in Table I, both algorithms deliver μE_h convergence within 13 iterations. For SiO, the OLS(1,1) method requires 33 iterations, whereas MIT(1,2) completes the task in only 12 cycles. Finally, for X $^1\Sigma_g^+$ C_2, OLS(1,1) fails for μE_h convergence, but MIT(1,2) is able to achieve this criterion in only 10 cycles, provided the model space $\mathbf{H}_{00}$ is increased to 750 determinants. Overall, in the best-behaved cases of our study, MIT(1,2) reduces the number of OLS(1,1) iterations required for μE_h convergence by 16% on average, but more importantly, for ground-state C_2 and SiO, MIT(1,2) averts the convergence difficulties of the Olsen single-vector approach. In summary, we find the OLS(1,1) and MIT(1,2) methods to be effective diagonalization

alternatives under the restricted conditions that a minimal number of expansion vectors is possible, limited energy convergence is acceptable, and the wave function is not of extreme multireference character. For tight convergence and multireference systems, MIT(1,2) is clearly more robust than OLS(1,1), but unless no other alternatives are feasible, neither method can be recommended under these circumstances.

Diagonalization performance relative to the $\mathbf{H}_{00}$ block size of the generalized Davidson preconditioner is presented in Table II. For the six molecular examples therein, expanding the dimension of $\mathbf{H}_{00}$ from 1 to 1000 reduces the number of iterations required for tight convergence by a substantial 15–24%, or 18% on average. Further expansion of the zeroth-order block size from 1000 to 4000 garners an additional mean reduction of 10%, giving an overall reduction of 27% with respect to the diagonal (DVD) preconditioner. Because the rate of recovery decays, the effects of adding more coupling in the preconditioner level off, and a point of diminishing return is soon reached, wherein the number of iterations saved does not pay for the extra CPU time spent in solving the expanded $\mathbf{H}_{00}$ linear system for $\boldsymbol{\delta}_0^k$ in eq. (14). This point of diminishing return is clearly dependent on the size of the CI problem. For computations involving up to 100 million determinants, we found it to generally be reached for block sizes less than 1000. Moreover, in exploratory computations on the effect of adding an intermediate coupling space in eq. (14), we found little improvement in convergence by the addition of $\mathbf{H}_{01}$ terms. In most of our computations, over half of the feasible benefit of GDVD preconditioning was achieved with an $\mathbf{H}_{00}$ block size of merely 400. In brief, we recommend the use of a GDVD preconditioner, with $\mathbf{H}_{00}$ block size of 400–1000, not only for obtaining substantial reductions in the number of required iterations for well-behaved systems but also for ensuring convergence in more taxing multireference applications.

Various methods for approximating diagonal elements in the preconditioner are compared in Table III for six molecules. The first salient feature of the data is that the use of average exchange integrals in a spin-coupling set (Method B) not only matches the performance obtained with exact diagonal elements, but in two of the examples actually reduces the number of iterations by 1. Therefore, Method B is able to maintain spin symmetry in the CI vector at each iteration in the diagonalization procedure while maintaining or even improving the rate of convergence. Among the three more approximate methods (C–E), the Evangelisti approach of eq. (22) (Method C) is clearly superior, suffering essentially no compromise in convergence compared to the exact method. In particular, in our test cases the number of iterations required by Method C never exceeds that required by either Method A or B by more than 1. For four of the molecules in Table III, Method D [eq. (21)] performs well, but it is inadequate for C_2 and SiO; on average, Method D requires 20% more iterations than the exact Method A. Finally, Method E, involving simple sums of orbital energies, is entirely unsatisfactory for all of the molecules, increasing the number of iterations anywhere from 50 to 200% over Method A. From our data we strongly recommend the use of average exchange integrals over spin-coupling sets (Method B), or the Evangelisti approach (Method C) if diagonal elements are to be computed on the fly. Both methods preserve spin symmetry in the expansion space.

Because the formation of one $\boldsymbol{\sigma}$ vector per root per iteration is the most CPU-intensive part of Davidson methods, simple iteration counts for various schemes, as presented in Tables I–III, are a primary indicator of efficiency. However, to ensure feasibility and to minimize the total wall time for any specific diagonalization problem, it is also necessary to weigh the costs of I/O operations and the external storage space requirements against the CPU expense of forming $\boldsymbol{\sigma}$ vectors. The implementation of collapse schemes is essential to reducing both I/O operations and storage requirements. The differences in I/O costs for various collapse schemes can be assessed by counting the minimum number of required operations (r_{IO}, w_{IO}) involving a (read,write) from/to disk of an entire vector of the dimension (N) of the CI problem, assuming only two storage buffers are available in memory whose lengths are some fraction of N. For our purposes we compare only single-root algorithms, started by diagonalizing a small $\mathbf{H}_{00}$ block, implemented with computation of diagonal Hamiltonian elements on the fly, and tested for convergence by an energy criterion and perhaps a residual vector norm from the previous iteration.

Let (n_c, n_b) denote the (lower, upper) limits of the subspace in the chosen collapse scheme, and let v be the current number of expansion vectors for a given iteration. Then an iteration of the Davidson method, with focus only on the I/O operation count, can be laid out as follows: (1) If $v = n_b$, expend $(r_{IO}, w_{IO}) = (2n_b n_c, 2n_c)$ to collapse each of the sets of $\mathbf{b}$ and $\boldsymbol{\sigma}$ vectors down to size n_c, assuming a straightforward quadratic contraction scheme

is invoked rather than a reordered, linear contraction algorithm;[54] (2) Expend $(r_{IO}, w_{IO}) = (2v, 2)$ to form and write the correction vector $\boldsymbol{\delta}$ to disk, while in the process rewriting the last **b** vector from the previous iteration to disk in normalized form (unnecessary immediately after a collapse); (3) Expend $(r_{IO}, w_{IO}) = (2v + 2, 1)$ to Schmidt orthogonalize the $\boldsymbol{\delta}$ vector against the existing expansion space and to write the result in unnormalized form to disk as a new **b** vector; (4) Perform an unspecified amount of I/O to form and write a new $\boldsymbol{\sigma}$ vector to disk, while simultaneously computing the next diagonal element of the **G** matrix; (5) Expend $(r_{IO}, w_{IO}) = (v + 1, 0)$ to compute the off-diagonal elements of **G** for the new **b** vector; (6) Diagonalize the expanded **G** matrix of dimension $v + 1$ to obtain new $\boldsymbol{\alpha}$ coefficients and new eigenvalue estimates. Test for convergence and perform another iteration if necessary.

The I/O requirements for step (4) do not depend on the collapse scheme and are thus immaterial for our comparisons. Moreover, the very first iteration actually requires only $r_{IO} + w_{IO} = 2$ full operations because the first **b** vector has nonzero elements only in the small, zeroth-order space. By summing up the I/O operations for successive iterations, we can arrive at the following formula for the total cost, $N_{IO} = r_{IO} + w_{IO}$, incurred if n iterations are required to converge the diagonalization procedure:

$$N_{IO} = \frac{m(5m+17)}{2} - 9 + q\left[\frac{n_b(5n_b+7) - n_c(5n_c+7)}{2}\right] + p\left(\frac{5p+10n_c+7}{2}\right) + (q+1-\delta_{p,0})\left[2n_c(n_b+1)-1\right] \quad (24)$$

where $m = \mathrm{Min}(n, n_b - 1)$, q is the greatest integer not exceeding $[\mathrm{Max}(n, n_b - 1) - n_b + 1](n_b - n_c)^{-1}$, $p = \mathrm{Max}(n, n_b - 1) - n_b + 1 - q(n_b - n_c)$, and $n > 1$. In essence, m is the number of startup iterations, q is the number of full collapse cycles of duration $n_b - n_c$, and p is the number of leftover iterations prior to convergence.

In Table IV, eq. (24) is applied to model the comparative I/O costs of the collapse schemes investigated in this article. It is worth emphasizing that using a linear contraction algorithm[54] rather than the straightforward quadratic approach would give reduced I/O costs, especially for larger (n_a, n_b) choices, but the data in Table IV are nonetheless instructive. For the full subspace method (∞, ∞), the I/O costs start to dramatically increase past 10 iterations. Indeed, in eq. (24) a large value of n_b gives $q = 0$, $p = 0$, and hence, $N_{IO} = (5n^2 + 17n - 18)/2$, showing that the I/O requirements for the full subspace method grow quadratically with the number of iterations. In contrast, in all collapse schemes there is only a linear dependence of I/O operations on n, and accordingly, past iteration 7, all such procedures involve a smaller N_{IO} count, in the general order $(1,2) < (1,3) < (1,4) < (2,4) < (2,3) < (\infty, \infty)$. To obtain benchmarks for typical tight diagonalization runs, note that at 10 iterations the I/O reduction compared to (∞, ∞) is (55, 53, 47, 26, 18%) for [(1,2), (1,3), (1,4), (2,4), (2,3)] collapse, and at 20 iterations these savings increase to (74, 72, 69, 55, 50%). As seen in the analysis of the data in Table I above, both the (2,4) and (2,3) schemes retain the convergence characteristics of the full subspace method, with the former slightly more efficient than the latter. Consideration of I/O loads enhances the preference for the (2,4) approach, because it requires about 10% fewer disk operations than (2,3). Increasing the number of collapse vectors (n_c) to 3 or greater increases N_{IO} of the (2,4) scheme over 50%,

TABLE IV.
Comparative I/O Costs (N_{IO}) of Various Collapse Schemes for the Davidson method.[a]

Iter	(∞, ∞)	(2,4)	(2,3)	(1,4)	(1,3)	(1,2)
2	18	18	18	18	18	18
3	39	39	49	39	36	34
4	65	74	80	59	52	50
5	96	95	111	75	70	66
6	132	130	142	96	86	82
7	173	151	173	116	104	98
8	219	186	204	132	120	114
9	270	207	235	153	138	130
10	326	242	266	173	154	146
11	387	263	297	189	172	162
12	453	298	328	210	188	178
13	524	319	359	230	206	194
14	600	354	390	246	222	210
15	681	375	421	267	240	226
16	767	410	452	287	256	242
17	858	431	483	303	274	258
18	954	466	514	324	290	274
19	1055	487	545	344	308	290
20	1161	522	576	360	324	306
21	1272	543	607	381	342	322
22	1388	578	638	401	358	338
23	1509	599	669	417	376	354
24	1635	634	700	438	392	370
25	1766	655	731	458	410	386

[a] Modeled according to eq. (24), and assuming a straightforward quadratic contraction scheme.

whereas increasing the expansion vector limit (n_b) to 5 results in I/O changes of only ±2%. Both of these observations argue for the (2,4) scheme as the optimal choice. However, if only three expansion and $\boldsymbol{\sigma}$ vectors can be stored, then the (2,3) alternative is almost as good, unless I/O costs dominate over $\boldsymbol{\sigma}$ vector formation, in which case the (1,3) approach will reduce the (2,3) I/O by over 40% at the expense of only 25% or fewer iterations for well-behaved cases.

Although the (2,3) approach requires more I/O operations than the (2,4) scheme, its performance is quite remarkable. In fact, the (2,3) approach opens up the possibility of retaining the convergence of the full subspace method when carrying out large-scale CI computations; if all six vectors can be stored in core memory, then the I/O costs associated with the (2,3) method are eliminated. As an example, consider the largest CI computation to date, a 10 billion determinant FCI energy.[18] The half a terabyte of memory necessary to store six vectors of that length currently exceeds that available on most computer clusters. However, several supercomputers at various national laboratories and supercomputing centers have up to eight times this storage capacity. For distributed computing, the (2,3) method does not incur any additional communication overhead relative to the Olsen single-vector iterator; the only nonlocal data operations are in the formation of the $\boldsymbol{\sigma}$ vector, which depends on the *current* $\boldsymbol{\delta}$ vector, not the total number of expansion vectors. Therefore, with the (2,3) method benchmark FCI property studies containing at least 10 billion determinants are now feasible.

Recommendations and Conclusions

The following conclusions are drawn from the current study.

1. In typical applications, placing coupling in the preconditioner via the generalized Davidson (GDVD) method accelerates convergence by about 20% for zeroth-order block sizes of 400–1000. The use of the GDVD preconditioner increases in importance, and may become critical for convergence, in multireference applications, when several roots are sought, or under severe constraints of the expansion space.
2. A point of diminishing return in expansion of the $\mathbf{H}_{00}$ block size in the GDVD preconditioner appears to be reached by 1000 determinants for CI problems less than 100 million, because the rate of convergence exhibits a pseudologarithmic dependence on the dimension of $\mathbf{H}_{00}$.
3. Invoking the (2,4) and (2,3) subspace collapse schemes achieves dramatic improvements in storage and I/O requirements at virtually no expense (<5%) in convergence rate. In typical applications, (1,4) and (1,3) collapse is expected to increase the number of iterations for tight convergence by less than 15 and 25%, respectively, the loses being reduced by using a GDVD preconditioner. The extreme (1,2) approach suffers from efficiency loses greater than 50%, and is prone to divergent behavior in multireference systems.
4. The Davidson (DVD) and Olsen (OLS) update vectors exhibit virtually identical convergence properties and no systematic differences in collapse behavior, except in the (1,2) extreme, where OLS may be preferred by up to 30%.
5. The use of least-squares extrapolation prior to subspace collapse has a neutral effect in typical applications, but in accord with the observations of Murray et al.,[28] it may yield substantial benefit if multiple roots are sought, the GDVD preconditioner is not employed, or the subspace size is severely limited.
6. The Olsen [OLS(1,1)] and Mitrushenkov [MIT(1,2)] one- and two-vector iterators, or similar alternatives,[55] are reasonable approaches for obtaining μE_h accuracy in very large CI problems, where storage and I/O requirements must be reduced in favor of additional CPU demands. However, these methods, particularly OLS(1,1), should be used with caution because they suffer severe convergence problems when there are near degeneracies in the eigenspectrum.
7. The diagonal element approximations involving averaged exchange integrals over a spin-coupling set and orbital energy differences referenced to the Hartree–Fock determinant [eq. (22)] both retain the convergence characteristics of the exact method while ensuring the maintenance of proper spin symmetry. The Evangelisti method[16] of eq. (22) is particularly well suited for efficient on-the-fly evaluation. The orbital energy sum approximation performs poorly and should be avoided.

For typical applications the (2,4) and (2,3) collapse schemes are highly recommended, and should be implemented with generalized Davidson preconditioners involving $\mathbf{H}_{00}$ blocks up to 1000, with

either Davidson or Olsen update vectors, and with one of the two preferred diagonal element approximations. The (2,3) scheme is particularly promising for applications where six vectors can be stored in memory, thus eliminating all I/O overhead without loss of convergence efficiency. The full subspace approach will provide no significant improvements in convergence over the (2,3) and (2,4) truncations and will carry substantially greater storage and I/O requirements. Collapsing the subspace to one vector brings slower rates of convergence, and carrying more than two vectors after a collapse has the same drawbacks as the full subspace method, i.e., greater storage and I/O burdens for no return. Least-squares extrapolation in typical applications not involving numerous roots is unnecessary.

Finally, although not considered in the present study, compression techniques and better initial guesses can improve the overall efficiency of the diagonalization procedure. Compression algorithms can be used to minimize I/O overhead by only storing to disk the coefficients larger than a given threshold[39, 47] or by employing other techniques.[45, 46] Improved initial guesses could be obtained by using a smaller CI wave function as an initial guess to a larger CI computation, for example, using a CISDTQ wave function as an initial guess for a CI computation including single through sixfold substitutions (CISDTQPH).

Although the recommended schemes were calibrated for symmetric (CI) matrices, the general techniques are extendable to nonsymmetric eigenvalue problems[56] found in electronic structure theory (e.g., EOM-CC, TD-DFT, and TD-HF). The recommended schemes have been used within the PSI3 package[51] to carry out FCI benchmarks on systems involving over a billion Slater determinants.[19, 20, 52, 53, 57, 58]

Appendix A: THE DAVIDSON–LIU ITERATIVE METHOD FOR THE LOWEST FEW EIGENVECTORS AND EIGENVALUES OF REAL, SYMMETRIC MATRICES (ADAPTED FROM REF. 27)

1. Select a set of L orthonormal guess vectors, at least one for each root desired, and place in the set $\{\mathbf{b}_i\}$.
2. Use a standard diagonalization method to solve the $L \times L$ eigenvalue problem

$$\mathbf{G}\boldsymbol{\alpha}^k = \rho^k \boldsymbol{\alpha}^k, \qquad k = 1, 2, \ldots, M \tag{25}$$

where

$$G_{ij} = (\mathbf{b}_i, \mathbf{H}\mathbf{b}_j) = (\mathbf{b}_i, \boldsymbol{\sigma}_j), \qquad 1 \le i, j \le L \tag{26}$$

and M is the number of roots of interest.

3. Form the correction vectors $\{\boldsymbol{\delta}^k\}$, $k = 1, 2, \ldots, M$, defined as

$$\delta_I^k = -\left(H_{II} - \rho^k\right)^{-1} r_I^k, \qquad I = 1, 2, \ldots, N \tag{27}$$

where

$$\mathbf{r}^k = \sum_{i=1}^{L} \alpha_i^k (\mathbf{H} - \rho^k)\mathbf{b}_i \tag{28}$$

and N is the number of determinants or configuration state functions.

4. Normalize $\{\boldsymbol{\delta}^k\}$.
5. Schmidt orthonormalize the first $\boldsymbol{\delta}^k$-vector against the set $\{\mathbf{b}_i\}$ and append the result to $\{\mathbf{b}_i\}$. Repeat this process for each of the other $M - 1$ correction vectors, neglecting any new vector whose norm after Schmidt orthogonalization is less than some threshold $T \sim 10^{-3}$. This results in the addition of m new $\mathbf{b}$ vectors, with $1 \le m \le M$.
6. Increase L by m and return to step 2.

Acknowledgments

M.L.L. thanks Edward Valeev for many helpful discussions. Sandia is a multiprogram laboratory operated by Sandia Corporation, a Lockheed Martin Company, for the United States Department of Energy under Contract DE-AC04-94AL85000.

References

1. Shavitt, I. Mol Phys 1998, 94, 3.
2. Čársky, P. In Encyclopedia of Computational Chemistry; Schleyer, P., Ed.; Wiley: New York, 1998; p. 485.
3. Sherrill, C. D.; Schaefer, H. F. In Advances in Quantum Chemistry; Löwdin, P.-O., Ed.; Academic Press: New York, 1999; p. 143, Vol. 34.
4. Roos, B. Chem Phys Lett 1972, 15, 153.
5. Handy, N. C. Chem Phys Lett 1980, 74, 280.
6. Siegbahn, P. E. M. J Chem Phys 1980, 72, 1647.
7. Siegbahn, P. E. M. Chem Phys Lett 1984, 109, 417.
8. Knowles, P. J.; Handy, N. C. Chem Phys Lett 1984, 111, 315.
9. Olsen, J.; Roos, B. O.; Jørgensen, P.; Aa. Jensen, H. J. J Chem Phys 1988, 89, 2185.
10. Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. Numerical Recipes in C; Cambridge Univ. Press: New York, 1992.
11. Parlett, B. N. The Symmetric Eigenvalue Problem; Prentice-Hall: Englewood Cliffs, NJ, 1980.

12. Davidson, E. R. J Comp Phys 1975, 17, 87.
13. Davidson, E. R. In Methods in Computational Molecular Physics; Diercksen, G. H. F.; Wilson, S., Eds.; NATO ASI Series C: Mathematical and Physical Sciences, vol. 113; D. Reidel: Boston, 1983; p. 95.
14. Davidson, E. R. Comput Phys Commun 1989, 53, 49.
15. Olsen, J.; Jørgensen, P.; Simons, J. Chem Phys Lett 1990, 169, 463.
16. Evangelisti, S.; Bendazzoli, G. L.; Ansaloni, R.; Durí, F.; Rossi, E. Chem Phys Lett 1996, 252, 437.
17. Ben-Amor, N.; Evangelisti, S.; Maynau, D.; Rossi, E. Chem Phys Lett 1998, 288, 348.
18. Rossi, E.; Bendazzoli, G. L.; Evangelisti, S.; Maynau, D. Chem Phys Lett 1999, 310, 530.
19. Tarczay, G.; Császár, A. G.; Leininger, M. L.; Klopper, W. Chem Phys Lett 2000, 322, 119.
20. Leininger, M. L.; Brown, S.; Barden, C.; Petraco, N. D. K.; Alien, W. D.; Schaefer, H. F. J Chem Phys, to appear.
21. Wood, D. M.; Zunger, A. J Phys A Math Gen 1985, 18, 1343.
22. Feler, M. G. J Comput Phys 1974, 14, 341.
23. Saad, Y. Numerical Methods for Large Eigenvalue Problems; Manchester University Press: Manchester, 1992.
24. Saad, Y.; Stathopoulos, A.; Wu, J. C. K.; Ogut, S. BIT 1996, 36, 563.
25. Morgan, R. B.; Scott, D. S. SIAM J Sci Stat Comput 1986, 7, 817.
26. Mitrushenkov, A. O. Chem Phys Lett 1994, 217, 559.
27. Liu, B. The simultaneous expansion method for the iterative solution of several of the lowest eigenvalues and corresponding eigenvectors of large real-symmetric matrices, Technical Report LBL-8158, Lawrence Berkeley Laboratory, University of California, Berkeley, 1978.
28. Murray, C. W.; Racine, S. C.; Davidson, E. R. J Comput Phys 1992, 103, 382.
29. van Lenthe, J. H.; Pulay, P. J Comput Chem 1990, 11, 1164.
30. Sleijpen, G. L. G.; van der Vorst, H. A. A generalized Jacobi–Davidson iteration method for linear eigenvalue problems, Technical Report Preprint 856, Dept. Math., Utrecht University, 1994.
31. Bofill, J. M.; Anglada, J. M. Chem Phys 1994, 183, 19.
32. van Dam, H. J. J.; van Lenthe, J. H.; Sleijpen, G. L. G.; van der Vorst, H. A. J Comput Chem 1996, 17, 267.
33. Bofill, J. M.; Anglada, J. M. Chem Phys 1994, 183, 19.
34. Bendazzoli, G. L.; Evangelisti, S. J Chem Phys 1993, 98, 3141.
35. Löwdin, P. O. In Perturbation Theory and Its Applications to Quantum Mechanics; Wilcox, C. H., Ed.; Wiley: New York, 1966.
36. Bendazzoli, G. L.; Evangelisti, S. Int J Quantum Chem Symp 1993, 27, 287.
37. Evangelisti, S.; Bendazzoli, G. L.; Gagliardi, L. Chem Phys 1994, 185, 47.
38. Evangelisti, S.; Bendazzoli, G. L.; Ansaloni, R.; Rossi, E. Chem Phys Lett 1995, 233, 353.
39. Ansaloni, R.; Bendazzoli, G. L.; Evangelisti, S.; Rossi, E. Comput Phys Commun 2000, 128, 496.
40. Malmqvist, P.-Å. In Lecture Notes in Quantum Chemistry: European Summer School in Quantum Chemistry; Roos, B. O., Ed.; Lecture Notes in Chemistry, vol. 58; Springer-Verlag: New York; 1992; p. 1.
41. Pople, J. A.; Krishnan, R.; Schlegel, H. B.; Binkley, J. S. Int J Quantum Chem Symp 1979, 13, 225.
42. Pauncz, R. Spin Eigenfunctions: Construction and Use; Plenum: New York, 1979.
43. Pauncz, R. The Symmetric Group in Quantum Chemistry; CRC Press: Boca Raton, FL, 1995.
44. Davidson, E. R.; Nitzsche, L. E.; McMurchie, L. E. Chem Phys Lett 1979, 62, 467.
45. Shepard, R. J Comput Chem 1990, 11, 45.
46. Dachsel, H.; Lischka, H. Theor Chim Acta 1995, 92, 339.
47. Rossi, E.; Bendazzoli, G. L.; Evangelisti, S. J Comput Chem 1998, 19, 658.
48. Dunning, T. H., Jr. J Chem Phys 1989, 90, 1007.
49. Dunning, T. H., Jr. J Chem Phys 1970, 53, 2823.
50. Dunning, T. H., Jr. J Chem Phys 1971, 55, 716.
51. Crawford, T. D.; Sherrill, C. D.; Valeev, E. F.; Fermann, J. T.; Leininger, M. L.; King, R. A.; Brown, S. T.; Janssen, C. L.; Seidl, E. T.; Yamaguchi, Y.; Allen, W. D.; Xie, Y.; Vacek, G.; Hamilton, T. P.; Kellogg, C. B.; Remington, R. B.; Schaefer, H. F., III. PSI 3.0; PSITECH, Inc.: Watkinsville, GA, 1999.
52. Leininger, M. L.; Sherrill, C. D.; Alien, W. D.; Schaefer, H. F. J Chem Phys 1998, 108, 6717.
53. Leininger, M. L.; Allen, W. D.; Schaefer, H. F.; Sherrill, C. D. J Chem Phys 2000, 112, 9213.
54. Shepard, R.; Shavitt, I.; Lischka, H., submitted for publication.
55. In a personal communication, S. Evangelisti recently reported that a single-vector iterator based on eq. (13), in which δ_k is not constructed to be orthogonal to x_k, performed as well or better than the OLS(1,1) approach. In limited tests, we found the two correction vectors to be comparable.
56. Hirao, K.; Nakatsuji, H. J Comp Phys 1982, 45, 246.
57. Sherrill, C. D.; Leininger, M. L.; Van Huis, T. J.; Schaefer, H. F. J Chem Phys 1998, 108, 1040.
58. Van Huis, T. J.; Leininger, M. L.; Sherrill, C. D.; Schaefer, H. F., III. Collect Czech Chem Commun 1998, 63, 1107.